CAN HUNCHES BE RATIONAL?

Gerd Gigerenzer[*] *and Henry Brighton*[**]

I. INTRODUCTION

Open any book on judgment and decision making and you will likely encounter two contrasting categories: rational and intuitive judgment. Rational judgment is defined by logical principles, such as the maximization of expected utility, Bayes' rule, or complex statistical prediction techniques. Despite the prevalence of such theories, people fail to adhere to these logical standards and instead rely on intuitive hunches, habits, and heuristics.

Books on the subject claim that short-cuts spring from our limited cognitive capacities and knowledge, which results in flawed reasoning and logical blunders. According to this view, mere hunches are inferior to logic and should be avoided unless time constraints and information costs leave no other choice. More information and more computational power, we are told, are always better. These conclusions tend to be presented as self-evident and obvious.

One glance at the world outside the confines of textbooks indicates that logical, reasoned decision making is not always superior to psychological intuition. Laypeople relying on mere hunches have outwitted financial analysts in predicting the stock market; simple heuristics have outperformed mutual funds and predicted the outcomes of 2003 Wimbledon tennis matches better than the official ATP expert rankings did (Borges et al. 1999; Serwe and Frings 2004; Törngren and Montgomery 2004). Zero-intelligence traders made as much profit as intelligent people did in experimental markets (Gode and Sunder 1993). Skilled athletes make better decisions when they have less time and information (Johnson and Raab 2003), and rely on heuristics for catching balls that require minimal information and ignore all variables relevant for computing the ball's trajectory (Shaffer and McBeath 2002). Limited memory capacities enable language learning, whereas larger capacities can prevent language acquisition in children as well as in neural networks (e.g., Elman 1993). Last but not least, *satisficers*, that is, people who search only for limited information and accept what is "good enough," report that they are more optimistic, self-confident and satisfied with their lives. In contrast, *maximizers*, people who opt for exhaustive search in order to find the absolutely best option, report depres-

---

[*] Gerd Gigerenzer is the director at the Max Planck Institute for Human Development in Berlin and the author of "Gut Feelings: The intelligence of the unconscious" (2007).

[**] Henry Brighton is a researcher at the Max Planck Institute for Human Development in Berlin.

Reprint: *Journal of Law, Economics & Policy, 4,* 155–175, 2007

sion, perfectionism, regret, and self-blame (Schwartz et al. 2002). These observations suffice to indicate a tension between the logical ideal that more information is always better and the psychological reality of intelligent hunches and heuristics.

This article draws on the research of the *adaptive toolbox* (Gigerenzer, Todd, and the ABC Research Group 1999; Gigerenzer and Selten 2001) to arrive at a better understanding of the nature and quality of hunches. A hunch is an intuitive judgment that appears quickly in our consciousness, and whose underlying reasons we are typically not aware of but nevertheless feel strongly enough to act upon (Gigerenzer 2007). This paper covers the following issues: First, many hunches are based on fast and frugal heuristics. That is, a hunch is the conscious product of an underlying, mostly unconscious process that is of heuristic (rather than analytic) nature. Second, we will show that simple heuristics that ignore information can be better—faster, more frugal, and more accurate—than complex strategies that use all available information. Third, we will clarify that heuristics are neither good nor bad, rational nor irrational, *per se*. Their performance depends on the structure of the environment. In order to understand when and why a simple heuristic works, one has to define the environmental structures it can exploit. The result of this program is a different conception of rationality, one that is not logical but ecological in nature. The concept of *ecological rationality* was independently developed by Gigerenzer et al. (1999) and Smith (2003). Hunches can be rational, but in a different and more efficient way than that suggested by logical rationality. Geographic profiling provides an illustrative example:

Seven armed robber locations in Victoria, Australia which were linked to the same offender. Where does the robber live?

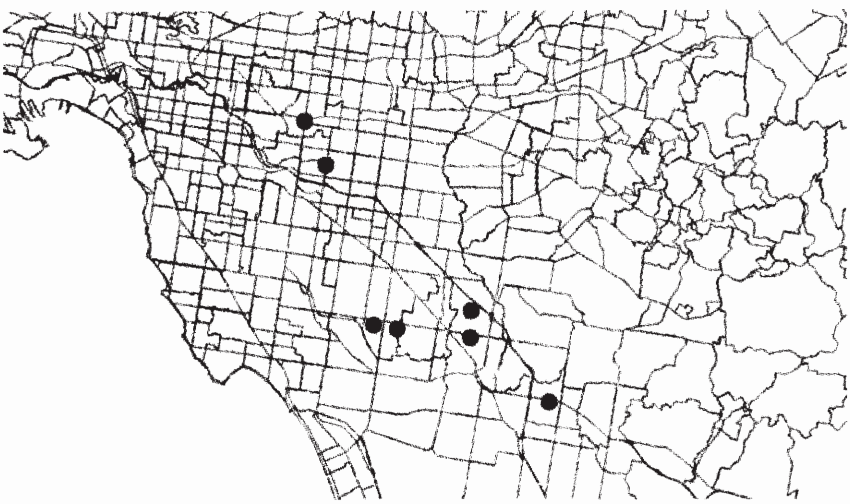

Figure 1

II. GEOGRAPHICAL PROFILING

A number of murders have been linked and the evidence points towards a serial killer (Figure 1). How can information about these crimes be used to focus police resources? Among several possible strategies, geographical profiling is one option. The locations of the crime scenes are used to predict the most likely location of the home of the offender. Geographical profiling can help locate perpetrators of serial crimes such as murder, burglary, arson, and armed robbery. By analyzing the relative locations of the crime scenes, resources can be put to their most efficient use by, for example, questioning known offenders residing within the locality predicted by the profiling strategy.

Given only the list of crime locations, we will consider two detectives with contrasting approaches: Detective *Satisficer* and Detective *Maximizer*. Satisficer uses a hunch to decide where the offender is most likely to live. Maximizer, on the other hand, turns to a state-of-the-art geographical profiling system. Unlike Satisficer, Maximizer will have received detailed and costly training in how to use the profiling software. Satisficer, who opts for the mental shortcut, is more likely to draw on past experience. Which detective will most effectively deploy police resources? Consider Figure 1, which depicts seven armed robbery locations in Victoria, Australia. Figure 2 illustrates the prediction of a commercial profiling system, CrimeStat (Levine 2000).

The predicted area of offender residence using CrimeStat. The true location of the offender is shown by the arrow.

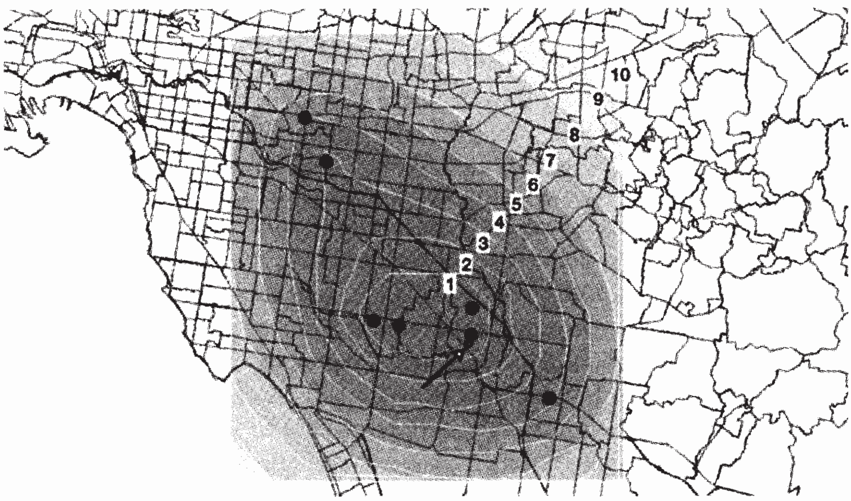

Figure 2

To make a prediction, CrimeStat performs a two-stage calculation. First, a distance decay function is applied to each crime location. This function assigns, for a single crime site, a likelihood score to every grid cell. Maximizer must choose the granularity of the grid, whereas the crime locations are real. For the second stage of the calculation, CrimeStat sums these individual likelihood scores to yield a final likelihood score for each grid cell. In particular, the first stage is carried out using the following function:

$$f(d_{ij}) = a \cdot e^{-c \cdot d_{ij}}$$

This function, where *i* represents the grid location of the crime location and *j* is an index over all other grid locations, maps the Euclidean distance between these locations, d_{ij}, to an offender residence likelihood score $f(d_{ij})$. The further from the crime location, the less likely that it is the offender's residence. Notice that the offender residence likelihood decreases exponentially with distance. The constants *c* and *a* are additional model parameters that alter the gradient of the exponential decay and the confidence of the likelihood scores, respectively. Given *n* crime locations, the second stage of the CrimeStat calculation sums the *n* likelihood scores that have been calculated for each grid cell. The result of this computation is a probability distribution covering the whole grid. The most likely offender residence is then predicted to be the locality with the highest likelihood score. The innermost region labelled "1" in Figure 2 is the predicted target area. This, in our example, represents the prediction of Maximizer. The actual residential location of the offender is shown by the arrow, and this location lies within the target region predicted by CrimeStat.

Now let us assume that Satisficer, who always acts on a hunch, uses a simple heuristic called the *circle heuristic*: take the two crime locations that are furthest apart, and then draw a circle passing through these two points. The circle heuristic predicts that the most likely offender location is at the centre of this circle. Figure 3 details the prediction of Satisficer. In this example, Satisficer's prediction is slightly outside the target area predicted by CrimeStat.

The predicted offender residence location predicted by the circle heuristic is marked by a square. The true offender residence is shown by an arrow.

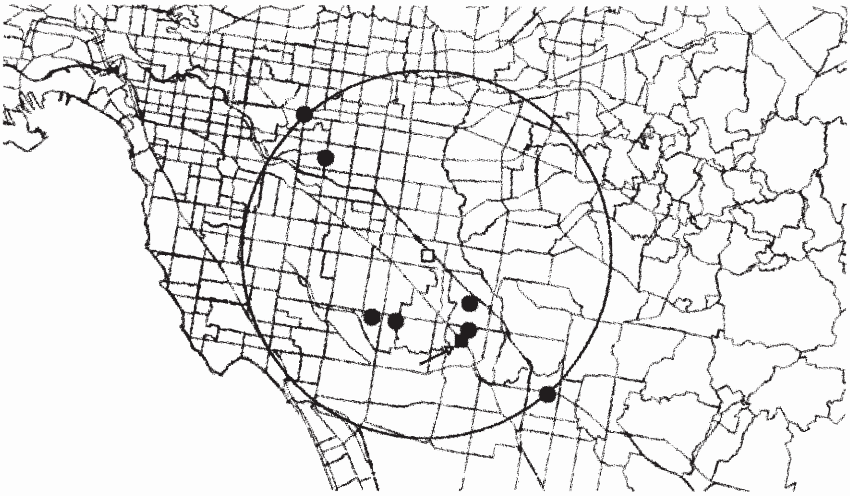

Figure 3

There are key differences between these two approaches. First, the computational costs of applying each strategy differ substantially. Second, Satisficer's prediction was only calculated on the basis of two of the crime locations (the two furthest from each other), whereas Maximizer's calculation used all the crime locations. In short, Maximizer's prediction required performing a complex computation taking into account all information. Satisficer's prediction drew on a fraction of the information and combined this information using a simple computation. Given the potential importance of an accurate prediction, the obvious answer is the most accurate. On the basis of the single example shown in Figures 1-3, one might conclude that a complex profiling system is more accurate on average than the circle heuristic. But is this true? When compared to complex and information hungry strategies, simple strategies that ignore information can perform just as well, or better, in predictive accuracy.

For example, Snook et al. (2005) compared the predictive accuracy of eleven geographical profiling strategies ranging from simple methods such as the circle heuristic to complex methods such as the probability distance strategy discussed above. The offenders considered were UK burglars who had committed a total of ten or more burglaries and, while resident at a single address, had committed between five and ten crimes. Snook et al. investigated the relationship between the number of crimes committed (while at a single address), strategy complexity (the computational cost of applying the strategy), and predictive accuracy (how far the prediction of the strategy deviates from the true residence of the offender). They discovered that in four out of six cases, a simple strategy (the circle heuristic) was the most accurate with a mean error distance of 8 km. The circle heuristic beat the best performing complex strategy by, on average, 1.25 km. In the two cases where the circle heuristic did not come out first, which where those with the largest number of crimes (nine and ten) committed, the most superior strategy was only 0.25 km more accurate than the simplest strategy. Snook et al. argued that the common assumption that strategy complexity is positively correlated with predictive accuracy is unfounded: the best average case accuracy was achieved by the circle heuristic. Only with an increasing number of crimes do the complex strategies begin to look worthy, but any advantage they achieve, Snook et al. argued, is insignificant: all strategies become more accurate when the number of crimes increases.

Demonstrating the accuracy of simple computational strategies over complex computational strategies is one thing, but how plausible is the claim that a real detective could use such a strategy? Can humans employ mental shortcuts similar to the circle heuristic? In another study, Snook, Taylor, and Bennett (2004) tested how well human subjects performed at predicting offender residences in comparison to CrimeStat. Three different subject groups were used. Two of these subject groups were exposed to one of two heuristics: the first group was shown the circle heuristic introduced above; the second group was introduced to the decay heuristic, which simply states that many offenders live near the location of their crimes. The third subject group, the control group, received no guidance on how to arrive at a prediction. Snook et al. found that, once again, the complex strategy did not perform significantly better than the two simple heuristics did. Subjects introduced to these heuristics rivaled the predictive accuracy of CrimeStat. Furthermore, they found that around half of human subjects in the control group, when asked, reported using a mental heuristic similar to those introduced: i.e., the subjects were basing their decisions on a hunch. Importantly, a significant proportion of subjects introduced to the heuristics improved their predictive accuracy as a result. The conclusion of Snook et al. (2004) was that people indeed employ simple and accurate heuristics, both with and without guidance.

Snook et al.'s work demonstrates that in spite of ignoring information, mental heuristics prove to be equally and sometimes more accurate than the complex and information hungry profiling strategy. This evidence suggests that Satisficer is, on average, likely to be at least as accurate as Maximizer.

III. ONE-REASON DECISION MAKING

The quality of hunches is examined in more detail in the following examples. The geographical profiling problem is all about finding the most probable location of an offender. Consider now a paired comparison task, such as which of two suspects committed a crime, which sport team will win the game, which of two schools will have a higher drop-out rake, or which of two stocks will yield a higher return. How does one construct a fast and frugal heuristic for a paired comparison task? One can use the same building blocks used in the fast and frugal geographical profiling heuristic: a search rule (consider the outermost crimes first), a stopping rule (stop when you have found two crimes), and a decision rule (predict the offender location to be the center of the circle passing through the two outermost crimes) (see Gigerenzer et al. 1999).

The "Take The Best" heuristic is designed for paired comparisons and uses these building blocks. It searches through cues, one by one. A search is terminated by a fast stopping rule: Stop when the first cue discriminates between the two alternatives. Finally, it uses a one-reason decision making rule: Only the cue that stops search determines the decision. The heuristic is called Take The Best because it relies on the best cue that discriminates and ignores the rest (Gigerenzer and Goldstein 1999). In general terms, the task is to predict which object, *a* or *b*, has the higher value on a criterion. There is a set of *N* objects and a set of *M* cues (1, 2, . . ., *i*, . . ., *M*). In the case of binary cues, cue values "1" and "0" indicate higher and lower criterion values, respectively. Take The Best can be characterized by the following building blocks:

(1) *Search rule*: Choose the cue with the highest validity and look up the cue values of the two objects.

(2) *Stopping rule*: If one object has a cue value of one ("1") and the other does not (i.e., "0" or unknown), then stop search and go on to Step 3. Otherwise exclude this cue and go back to Step 1. If no cues are left, guess.

(3) *Decision rule*: Predict that the object with the cue value of one ("1") has the higher value on the criterion.

The validity v_i of a cue i (Step 1) is defined as

$$v_i = \frac{R_i}{R_i + W_i}$$

where R_i is the number of correct inferences, and W_i is the number of incorrect inferences based on cue i alone. $R_i + W_i$ equals the number of cases where one object has the value "1" and the other does not.

The "Minimalist" heuristic is a close relative of Take The Best, and differs only in the search rule. This heuristic simply picks cues in random order:

(1) *Search rule*: Draw a cue randomly (without replacement) and look up the cue values of the two objects.

Each of the three building blocks of Take The Best (and even more, in the Minimalist heuristic) bets on the power of simplicity. The search rule looks up cues in the order of their validities. To order cues according to v_i is fast and frugal but not "optimal," because this order ignores dependencies between cues. Like the search rule, the stopping rule does not employ optimization calculations either. No attempt is made to calculate the point where the costs of further search will exceed its benefits. The decision rule violates the ideal of compensation embodied in all standard theories of rational choice: to arrive at a decision by weighing and adding. The decision rule bases the prediction only on the best cue that differentiates between the two alternatives, that is, it uses one-reason decision making.

The two heuristics belong to a class of heuristics that employ *one-reason decision making* (Gigerenzer 2004). The term refers to the fact that the decision is based on only one cue (the decision rule), while search can go through several cues (the search rule). Empirical evidence shows that intuitive judgments are often based on one reason only, both in experimental situations with low stakes (Bröder and Schiffer 2003; Rieskamp and Hoffrage 1999; Shepard 1964), and in situations with high stakes, such as when parents choose primary health care for their sick child during nighttime (Scott 2002). One-reason decision making seems to be mostly unconscious, and is a possible candidate for the process underlying some forms of hunches.

Because rational judgment is often defined by logical principles, intuitions that are based on one reason are considered a form of human irrationality, a regrettable byproduct of our "cognitive limitations." This interpretation is still characteristic for behavioral economics, and can be found in the behavioral law and economics literature as well. This normative claim is made on logical grounds, but, as far as we know, has never been tested. The following tests the validity of one-reason decision making in the form of Take The Best method when faced with complex real-world problems rather than logical textbook problems.

A description of the twenty prediction problems used in the competition.		
All data sources are listed in Czerlinski et al. (1999). For each problem, we specify the criterion and give a sample of the cues for predicting the criterion. The cues are either binary or were dichotomized by a median split. The raw data is available via world wide web at http://www.mpib-berlin.mpg.de/abc/.		
Psychology	*Attractiveness of men*	Predict average attractiveness ratings of thirty-two famous men based on the average likeability ratings of each man, the percent of subjects who recognized the man's name (subjects saw only the name, no photos), and whether the man was American. Based on data from 115 male and 131 female Germans, aged 17-66 years.
	Attractiveness of women	Predict average attractiveness ratings of thirty famous women based on the subjects' average likeability ratings of each woman, the percent of subjects who recognized the woman's name (subjects saw only the name, no photos), and whether the woman was American. Based on data from 115 male and 131 female Germans, aged 17-66 years.
Sociology	*High school dropout rate*	Predict drop-out rate of the fifty-seven Chicago public high schools, given the percentage of low-income students, percentage of non-White students, average SAT scores, etc.
	Homelessness	Predict the rate of homelessness in fifty U.S. cities given the average temperature, unemployment rate, percent of inhabitants with incomes below the poverty line, the vacancy rate, whether the city has rent control, and the percent public housing.
Demography	*Mortality*	Predict the mortality rate in twenty U.S. cities given the average January temperature, HC pollution level, the percentage of non-White residents, etc.
	City population	Predict populations of the eighty-three German cities with at least 100,000 inhabitants based on whether each city has a soccer team, university, intercity train line, exposition site, etc.
Economics	*House price*	Predict the selling price of twenty-two houses in Erie, PA, based on current property taxes, number of bathrooms, number of bedrooms, lot size, total living space, garage space, age of house, etc.
	Land rent	Predict the rent per acre paid in fifty-eight counties in Minnesota (in 1977 for agricultural land planted in alfalfa) based on the average rent for all tillable land, density of dairy cows, proportion of pasture land, and whether liming is required to grow alfalfa on the land. (Alfalfa is often fed to dairy cows.)
	Professors' salaries	Predict the salaries of fifty-one professors at a Midwestern college given gender, rank, number of years in current rank, the highest degree earned, and number of years since highest degree earned.
Transportation	*Car accidents*	Predict the accident rate per million vehicle miles for thirty-seven segments of highway, using the segment's length, average traffic count, percent of truck volume, speed limit, number of lanes, lane width, shoulder width, number of intersections, etc. for Minnesota in 1973.
	Fuel consumption	Predict the average motor fuel consumption per person for each of the forty-eight contiguous United States using the population of the state, number of licensed drivers, fuel tax, per capita income, miles of primary highways, etc.
Health	*Obesity at age 18*	Predict body fat percentages at age 18 of forty-six children based on measurements from ages 2 to 18. The body measurements include height, weight, leg circumference, and strength. (Based on the longitudinal monitoring of the Berkeley Guidance Study.)
	Body fat	Predict percentage of body fat determined by underwater weighing (a more accurate measure of body fat) using various body circumference measurements (which are more often used because they are more convenient measures than underwater weighing) for 218 men.
Biology	*Fish fertility*	Predict the number of eggs in 395 female Arctic charr based on the fish's weight, its age, and the average weight of its eggs.
	Mammals' sleep	Predict the average amount of time thirty-five species of mammals sleep, based on brain weight, body weight, life span, gestation time, and predation and danger indices.
	Cow manure	Predict the amount of oxygen absorbed by dairy wastes given the biological oxygen demand, chemical oxygen demand, total Kjedahl nitrogen, total solids, and total volatile solids for fourteen trials.
Environmental Science	*Biodiversity*	Predict the number of species on twenty-six Galapagos islands, given the area, elevation, distance to the nearest island, area of the nearest island, distance from the coast, etc.
	Rainfall from cloud seeding	Predict the amount of rainfall on twenty-four days in Coral Gables, Florida, given the types of clouds, the percent of cloud cover, whether the clouds were seeded, number of days since the first day of the experiment, etc.
	Oxidant in Los Angeles	Predict the amount of oxidant in Los Angeles for seventeen days given the day's windspeed, temperature, humidity, and insolation (a measure of the amount of sunlight). Data provided by the Los Angeles Pollution Control District.
	Ozone in San Francisco	Predict the amount of ozone in San Francisco on eleven occasions based on the year, average winter precipitation for the last two years, and ozone level in San Jose, at the southern end of the Bay.

Table 1

A. *First Competition*

Table 1 lists twenty demographic, economic, psychological, biological, and environmental prediction problems studied by Czerlinski, Gigerenzer, and Goldstein (1999). The number of cues (predictors) varied between three and nineteen, and these were binary or dichotomized at the median. In each case, the task was to predict which of two objects scores higher on a criterion. For instance, one task was to predict which of Chicago public high schools *a* and *b* has the higher dropout rate. The cues included attendance rates of the students, socio-economic and ethnic compositions of the student bodies, sizes of the classes, parental participation rates, and the scores of the students on various standardized tests. The search rule of Take The Best first looks up the information concerning attendance rate. If school *a* has a high attendance rate, but school *b* does not, search is stopped and the inference is made that high school *a* will have the larger dropout rate. If the condition of the stopping rule is not met, then search is continued for the cue with the second highest validity, and so on. The Minimalist, in contrast, looks up cues in random order.

Czerlinski et al. (1999) compared the performance of the two heuristics with that of multiple regression, and with a simple tallying heuristic that does not calculate the beta weights but only uses unit weights of +1 or -1 (Dawes 1979). There were two tasks: data fitting and prediction. The difference between data fitting and prediction is of great importance and can be understood in analogy to hindsight versus foresight. In hindsight, one already knows what happened, and the task is to construct an explanation post hoc; in foresight, one does not know what will happen, and has to make a true prediction from our theory. Similarly, in data fitting, one already knows the data and fits the parameters of a model post hoc so that they achieve a maximum fit. Prediction, in contrast, is the true test of a theory of human judgment, its moment of truth.

In data fitting, each of the four models had the complete data of each of the twenty problems available, and tried to fit this data. In prediction, each model learned its parameters from half of the objects (training set), and was tested on the other half (test set), a procedure known as cross-validation. Multiple regression, for instance, estimated in the training set its beta weights, whereas Take The Best estimated the order of cues. Figure 4 shows that the two linear models used on average 7.7 cues (exhaustive search), whereas Take The Best and the Minimalist only looked up 2.4 and 2.2 cues, respectively (limited search). Both heuristics were quite frugal, but how accurate were they? The usual assumption is that frugality comes at the price of lower accuracy. Is that true?

A competition between heuristics and multiple regression (Czerlinski et al. 1999). Accuracy is measured for data fitting (hindsight) and prediction (foresight). Results are averaged across twenty different real-world prediction tasks (Table 1). For each of the twenty problems and each of the four strategies, the 95 percent confidence intervals were ≤ 0.42 percentage points. Take The Best and Minimalist are heuristics that practice one-reason decision making, tallying attends to all reasons but ignores weights, and multiple regression uses all information available, calculates its optimal weights, and combines all cues linearly.

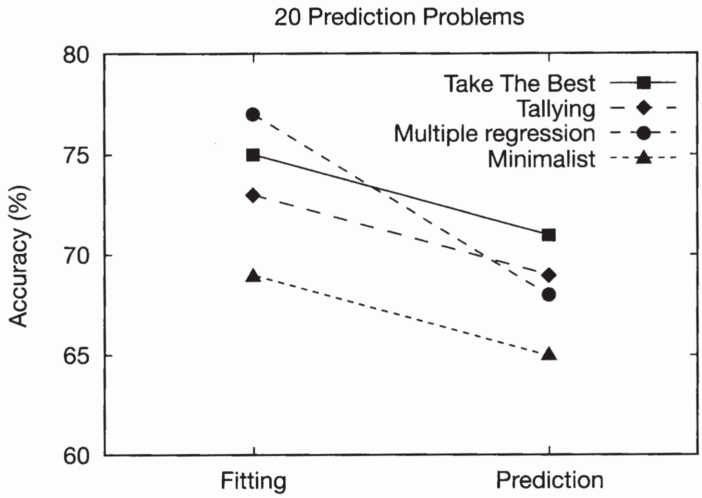

Figure 4

Figure 4 shows the results across all twenty problems. In data fitting (hindsight), multiple regression fitted the data best, followed by Take The Best and the unit-weight linear model. It is remarkable how close Take The Best came to multiple regression, and that it actually was more accurate than the tallying heuristic. In prediction (foresight), however, regression was no longer ahead. The predictive accuracy of tallying outperformed that of multiple regression despite its optimal beta weights. This apparently paradoxical effect—that the "optimal" weights (beta coefficients) are not better than "improper" unit weights of +1 or -1—has been reported earlier (e.g., Dawes 1979; Einhorn and Hogarth 1975). Yet since then, this result has been successfully repressed in the collective memory of decision theory (Hogarth 2005). On average, tallying made more accurate predictions for the twenty complex problems.

The Minimalist heuristic ignores information about the quantitative weights of cues (just as tallying does) and relies on one-reason decision making (just as Take The Best does). The performance of the Minimalist is substantially lower in both fitting and prediction, indicating that ignoring both weights and cues is too much. For these twenty complex problems, ignoring either cues or their weights is beneficial, but not both.

We surely would be surprised to find that the frugal model produced results comparable to the high end model, the model with ostensibly "better inputs." But how should we react when we find that the frugal model produces results not merely adequate, but demonstrably superior? The performance of the Take the Best Model, in short, presents to us a profoundly counterintuitive result. The predictive accuracy of Take the Best was, on average, higher than that of multiple regression and the other competitors. This may appear paradoxical because multiple regression processed all the information that Take The Best did and more. More remarkable still, regression used complex computational algorithms that require the use of a computer, whereas Take The Best can be done mentally.

And yet, this is not to say that less is necessarily more. The Minimalist heuristic ignores information about the quantitative weights of cues (just as tallying does) and relies on one-reason decision making (just as Take The Best does). The performance of the Minimalist is substantially lower in both fitting and prediction, indicating that ignoring both weights and cues is too much. For these twenty complex problems, ignoring either cues or their weights is beneficial, but not both.

B. *Policy implications*

Knowing which strategy has the highest predictive accuracy can help policy-makers determine how to weigh the various factors in a complex policy problem. For example, Take The Best regarded attendance rate, writing test score, and social science test score as the most valid cues for high school dropout rates, in that order. In contrast, linear regression's top three predictors were percentage of Hispanic students, percentage of students with limited English, and percentage of Black students. The different models each employ a different strategy. Each strategy suggests a different course of action to the policy-maker seeking to lower dropout rates. While the Take The Best analysis would recommend getting students to attend class and teaching them the basics more thoroughly, a regression user would advocate helping minorities assimilate and supporting "English as a Second Language" (ESL) programs.

Again we face the question: "How can one reason be as good as or better than many?" Take The Best strikes a balance between the dangers of overfitting (that is, extracting too much information—noise—from the training set, as multiple regression did) and underfitting (extracting too little information, as the Minimalist did). Generally, a model *A* overfits the training data if there exists an alternative model *B*, such that the accuracy of *A* is higher than or equal to B in the training set, but lower in the test set. Before we take up the question of why and when one-reason decision making works in more detail, we will first send Take The Best back into the ring for another round of tests. This time, the heuristic will meet the reputed world champions among complex strategies.

C. *Second Competition With Heavyweight Contestants*

The problem of overfitting is a fundamental concern for those interested in machine learning, where sophisticated algorithms are designed with a view to maximizing predictive accuracy (e.g., Mitchell 1997). Those interested in machine learning seek theories of how learning tasks can be accomplished, both from a psychological and an engineering perspective. Machine learning algorithms, in comparison to the competitors discussed above, are typically more complex: they are designed to achieve a high degree of predictive accuracy over a wider range of problems. The product of many years of research in computational learning, these complex algorithms have been tested and applied in the context of many different disciplines: artificial intelligence, evolutionary psychology, etc. These complex algorithms are, in short, true heavyweight competitors, durable and tested, with a strong predictive power. How, then, does the predictive accuracy of Take The Best compare to the predictive accuracy of some standard machine learning algorithms?

We compared Take The Best to three heavyweight competitors, all of which, in a number of guises, have been proposed as models of human decision making (e.g., Chater et al. 2003). The competitors are: (1) the decision tree induction algorithm C4.5 (Quinlan 1993), (2) a feed-forward neural network trained using backpropagation (we will refer to this model as BackProp; Rumelhart, McClelland, and the PDF Research Group 1986), and (3) the nearest neighbor decision rule (referred to below as 1-NN; Cover and Hart 1967).

C4.5. For example, C4.5 constructs a decision tree that represents a system of rules by first growing a tree using the information theoretic measure of entropy. It then prunes back this tree, which results in the rules becoming less specific, in an attempt to avoid overfitting noisy information.

D. *Neural Network*

A neural network addresses the same problem very differently. A neural network encodes an abstract representation of past solutions to a problem using a network of weighted connections between artificial neurons. Several thousand weight updates are required for the network to learn from the past examples, but once it has, trained networks are often very robust against overfitting as they rarely represent hard and fast rules.

E. *Nearest Neighbor Decision Rule*

The third competitor, the nearest neighbor decision rule (1-NN), is an example of an exemplar model in that it stores all presented training examples in a memory. When a prediction is required for a novel example of the problem (such as a previously unseen paired comparison), the most similar previously encountered example is retrieved from memory. Similar to solving a problem by analogy, the exemplar model uses a similar previously encountered example of the problem to propose a solution to the new problem.

A competition between Take The Best and three heavy-weight machine learning algorithms, C4.5, BackProp, and NN-1, over four diverse environments taken from the twenty shown in Table 1. (In the lower right panel, C4.5 and BackProp overlap.)

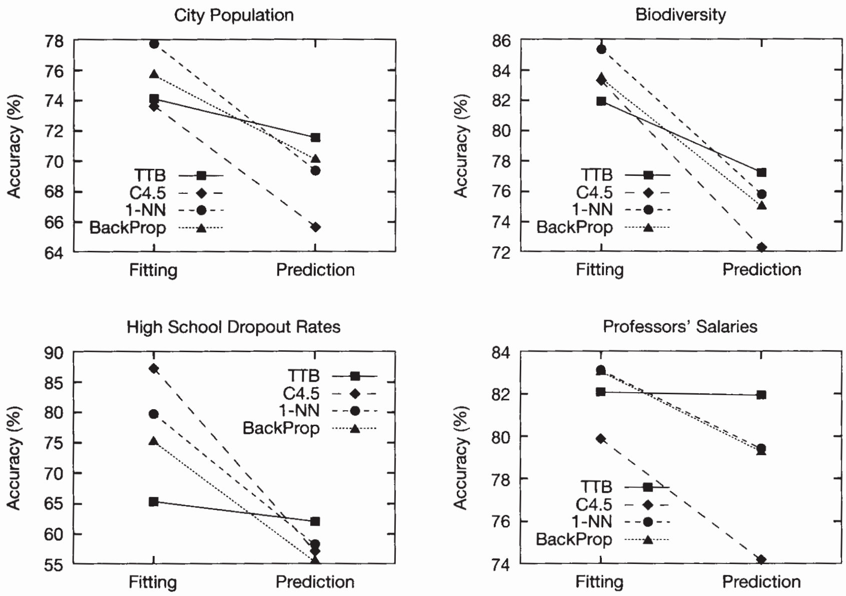

Figure 5

Figure 5 shows how well these three competitors performed in comparison to Take The Best in four tasks which represent the twenty environments contained in the four major complex problems previously discussed (i.e., city populations, biodiversity, high school dropout rates, and professors' salaries.) In all four complex problems, we observed a similar pattern. If the task was data fitting, the complex strategies were substantially better than Take The Best. Yet when these strategies were put to the more important task of prediction, their performance had a steep decline. That is, they overfitted substantially, despite the various ways in which they have been designed to tackle this problem.

In contrast, Take The Best showed a relatively small propensity to overfit and achieved the highest proportion of correct predictions in each of the four complex problems. Predictive accuracy is not only the task faced by detectives, but also the relevant criterion with which to judge competing models of cognition (Pitt, Myung, and Zhang 2002). The fact that heuristics based on one reason can outperform the most sophisticated complex strategies is an important result that has not yet been demonstrated before.

This heavyweight competition offers further evidence that a heuristic that ignores information can achieve higher predictive accuracy than complex strategies that take much more information into account. Both the neural network model and the exemplar model use all cue values to inform their decision: they weigh and add each cue value to yield a final decision. Furthermore, the results in Figure 5 also show the decision tree algorithm C4.5 to be suffering from a tendency to overfit. This tells us that the decision trees constructed by C4.5 are focusing on irrelevant information. The trees are too specific in what they consider to be informative, which means that too many cues are being considered when making a prediction. The three sophisticated competitors may well identify useful patterns in the data, but this information is likely to be rendered unreliable because all three competitors also identify and act on irrelevant information. On the other hand, Take The Best, the simpler decision strategy, focuses on the useful information and acts on this information alone.

IV. ECOLOGICAL RATIONALITY

How can one reason be better than many? There are two answers. One is that the robustness of simple heuristics protects against overfitting. In a situation where there is uncertainty—and there is, for instance, a lot of uncertainty in predicting dropout rates—only part of the information obtainable today will be of predictive value for the future (Geman, Bienenstock, and Doursat 1992). If one records the temperature of each day of this year in a city, one can find a mathematical equation with sufficiently complex exponential terms that represents the jagged temperature curve almost perfectly.

However, this equation may not be the best predictor of next year's temperature; a simpler curve that ignores much of this year's measures may do better. In other words, only part of the information available in one situation generalizes to another. To make good inferences or predictions under uncertainty, *one has to ignore part of the information available*. The art is to find the part that generalizes. Since Take The Best relies only on the best cue, its chances of ignoring less reliable information are good.

Consider two diagnostic systems, one with more adjustable parameters (e.g., predictors) and one with only a subset of these, that is, with less.

Both systems fit a given body of data (e.g., a sample of patients) equally well. When making predictions about a new sample, the general result is that the simpler system will make more accurate predictions than the system with more parameters will. This form of less-is-more has been mathematically proven for specific situations (see Akaike 1973; Forster and Sober 1994; Geman et al. 1992). With a sufficient number of parameters, one can always fit a sample of observations. In general, the more unpredictable the situation is, the more information should be ignored. The art of good decision making is to focus on that part of the information that generalizes and to ignore the rest. This is what a good hunch does.

Heuristics thrive on particular structures of environments. The left side shows an environment that consists of cues whose weights are noncompensatory (e.g., 1/2, 1/4, 1/8, and so on). In this environment, no weighted linear model can achieve a higher fit than the faster and more frugal Take The Best. The right side shows a compensatory environment, where linear models will have the advantage (Martignon and Hoffrage 1999).

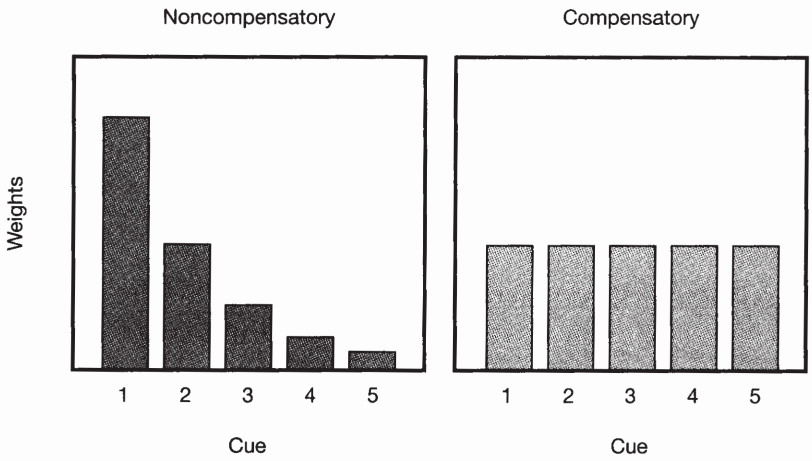

Figure 6

The second answer is the concept of *ecological rationality*, meaning that the match of a heuristic with the structure of environments limits both overfitting and underfitting. There are several structures that one-reason decision making can thrive on (Martignon and Hoffrage 2002). Recall that Take The Best is a noncompensatory strategy: it relies on one cue, and even if all other cues point in the opposite direction, they cannot compensate. One of several structures that Take The Best can exploit is noncompensatory information. Figure 6 shows examples for noncompensatory and compensatory environments. For instance, binary cues with weights that decrease exponentially, such as 1, 1/2, 1/4, 1/8, and so on, are noncompensatory—the sum of all cue weights to the right of a cue can never be larger than its own weight. When the environment has the same noncompensatory structure as Take The Best, one can prove that no linear model, including multiple regression, can obtain a better fit than does the faster and more frugal Take The Best (Martignon and Hoffrage 2002). The right side of Figure 6 shows an environmental structure where Take The Best is less successful but which the tallying heuristic can exploit. In the extreme case shown with equal weights, it is obvious that no sophisticated linear model can outperform the simpler version.

V. THE ADAPTIVE TOOLBOX

Modern statistical technology has become an attractive alternative to intuitive judgment, and even informed intuitions are seen as inferior to complex computational strategies. Blind trust in complexity and distrust of informed intuition, however, needs to be replaced by a systematic study of the quality of both. The first step is to explicate the mechanisms that produce hunches. This is not easy, since people, including experts, do not always know how they arrive at a given judgment. We have argued that hunches, at least one class of them, can be explicated in terms of fast and frugal heuristics. That first step leads to the construction of models of heuristics and, in a second step, allows these to be tested in comparison with strategies that use more information and require more computation.

Contrary to the wisdom implicit in most of decision theory, the results we reported indicate that heuristics that base their decision (the hunch) on only one reason are often as accurate as, if not more so, than the most sophisticated statistical strategies available today. Note that our tests assumed heuristics used by people who are not totally ignorant but somewhat knowledgeable, that is, they had a learning phase to estimate what the most important cues are. Our tests did not, however, assume that people know how to weigh cues quantitatively, calculate the dependencies between cues, and integrate these into a final judgment. A number of studies have correctly concluded that people fail to do these computations, but incorrectly made the further inference that this is a sign of mental deficiency. As we have shown, a heuristic that ignores dependencies between cues can actually achieve better results than can strategies that are able to compute dependencies.

What are the consequences of this research for training experts in making good predictions? The way to go, in our view, is to systematically perform research on heuristics for the problem at hand, and to train experts in using, checking, and updating these. This is an alternative to both the traditional "rational choice" training in expected utility maximization, and the replacement of human experts by expensive statistical forecasting technology. In high-technology and high-stakes areas such as medical diagnostics, the systematic teaching of fast and frugal heuristics to doctors is already under discussion (Elwyn et al 2001; Green and Mehr 1997; Naylor 2001). Given the time pressure and uncertainties of diagnosis, physicians in fact already rely on heuristics, but they do not always admit it, in fear of legal suits. As a result, physicians' hunches tend to be inconsistent, varying from physicians to physician and from teaching hospital to teaching hospital (Gigerenzer 2002). Likewise, police officers, fearful of judicial rebuke, conceal the nature of their thought processes—why they stopped one suspect and not another, why they frisked him, but not her. Their testimony in suppression herrings is larded with "reasons," many of which in fact played little or no role in their decision to make a stop and frisk. Heuristics need to be discussed and evaluated openly, just as complex computational strategies should be checked as to how successful they are. Decision theory and its applications need to move away from the emotional attachment to logical ideals of rationality, and to acquire a more competitive, empirical spirit.

Can hunches be rational? In this chapter we have argued that they can. We have reviewed evidence that heuristics that rely on one good reason can be as accurate as or better than a complex analysis that weighs the pros and cons of multiple factors. The wisdom of a hunch is precisely that it bets on what is important and ignore the rest. Oddly, the current American legal regime insists that police officers cite legions of "objective" data in a suppression hearing, when the fact is that, in many circumstances, an officer who acted on less information will achieve greater success than an officer who tabulated dozens of factors in his mind before acting, if he ever acted at all. We may even speculate that some of the most successful police officers are, to recall the terminology in the beginning of the chapter, "satsificers" when they stop a suspect (acting on little information). By the time they take the witness stand, however, they have will have remade themselves into "maximizers," detailing to an attentive, and usually credulous judge, the myriad of factors that supposedly spurred them to act.

REFERENCES

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory*, ed. B. N. Petrov and F. Csáki, 267-81. Budapest: Akademiai Kiado.

Borges, B., D. G. Goldstein, A. Ortmann, and G. Gigerenzer. 1999. Can ignorance beat the stock market? In *Simple heuristics that make us smart*, G. Gigerenzer, P. M. Todd, and the ABC Research Group, 59-72. New York: Oxford University Press.

Bröder, A., and S. Schiffer. 2003. Take The Best versus simultaneous feature matching: probabilistic inferences from memory and effects of representation format. *Journal of Experimental Psychology: General* 132: 277-93.

Chater, N., M. Oaksford, R. Nakisa, and M. Redington. 2003. Fast, frugal, and rational: how rational norms explain behavior. *Organizational Behavior and Human Decision Processes* 90: 63-86.

Cover, T. M., and P. T. Hart. 1967. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory* 13: 21-27.

Czerlinski, J., G. Gigerenzer, and D. G. Goldstein. 1999. How good are simple heuristics? In *Simple heuristics that make us smart*, G. Gigerenzer, P. M. Todd, and the ABC Research Group, 97-118. New York: Oxford University Press.

Dawes, R. M. 1979. The robust beauty of improper linear models in decision making. *American Psychologist* 34: 571-82.

Einhorn, H. J., and R. M. Hogarth. 1975. Unit weighting schemes for decision making. *Organizational Behavior and Human Performance* 13: 171-92.

Elman, J. L. 1993. Learning and development in neural networks: the importance of starting small. *Cognition* 48: 71-99.

Elwyn, G., A. Edwards, M. Eccles, and D. Rovner. 2001. Decision analysis in patient care. *The Lancet* 358: 571-74.

Forster, M., and E. Sober. 1994. How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. *British Journal of Philosophical Science* 45: 1-35.

Geman, S. E., E. Bienenstock, and R. Doursat. 1992. Neural networks and the bias/variance dilemma. *Neural Computation* 4: 1-58.

Gigerenzer, G. 1999. Betting on one good reason: the Take The Best heuristic. In *Simple heuristics that make us smart*, G. Gigerenzer, P. M. Todd, and the ABC Research Group, 75-95. New York: Oxford University Press.

—— 2002. *Calculated risks: how to know when numbers deceive you*. New York: Simon & Schuster. (United Kingdom version: *Reckoning with risk: learning to live with uncertainty*. London: Penguin Books.)

—— 2004. Fast and frugal heuristics: the tools of bounded rationality. In *Blackwell handbook of judgment and decision making*, ed. D. J. Koehler and N. Harvey, 62-88. Oxford, United Kingdom: Blackwell.

—— 2007. *Gut feelings: the intelligence of the unconscious*. New York: Viking. (United Kingdom version: London: Allen Lane/Penguin.)

Gigerenzer, G., and D. G. Goldstein. 1996. Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review* 103: 684-704.

Gigerenzer, G., and R. Selten, eds. 2001. *Bounded rationality: the adaptive toolbox*. Cambridge, Massachusetts: MIT Press.

Gigerenzer, G., P. M. Todd, and the ABC Research Group. 1999. *Simple heuristics that make us smart*. New York: Oxford University Press.

Gode, D. K., and S. Sunder. 1993. Allocative efficiency of markets with zero-intelligence traders: market as a partial substitute for individual rationality. *Journal of Political Economy* 101: 119-37.

Green, L., and D. R. Mehr. 1997. What alters physicians' decisions to admit to the coronary care unit? *The Journal of Family Practice* 45: 219-26.

Hogarth, R. M. (In press.) The bumpy road to enlightenment: a history of some simple heuristics. In *Ecological rationality: intelligence in the world*, ed. G. Gigerenzer, P. M. Todd, and the ABC Research Group. New York: Oxford University Press.

Johnson, J. G., and M. Raab. 2003. Take the first: option generation and resulting choices. *Organizational Behavior and Human Decision Processes* 91: 215-29.

Levine, N. 2000. *CrimeStat: a spatial statistics program for the analysis of crime incident locations (Version 1.1)*. Washington, D.C.: Ned Levine & Associates/National Institute of Justice.

Martignon, L., and U. Hoffrage. 1999. Why does one-reason decision making work? A case study in ecological rationality. In *Simple heuristics that make us smart*, G. Gigerenzer, P. M. Todd, and the ABC Research Group, 119-40. New York: Oxford University Press.

—— 2002. Fast, frugal and fit: Lexicographic heuristics for paired comparison. *Theory and Decision* 52: 29-71.

Mitchell, T. M. 1997. *Machine Learning*. New York: McGraw-Hill.

Naylor, C. D. 2001. Clinical decisions: from art to science and back again. *The Lancet* 358: 523-24.

Pitt, M. A., J. Myung, and S. Zhang. 2002. Toward a method of selecting among computational models of cognition. *Psychological Review* 109: 472-91.

Quinlan, J. R. 1993. *C4.5: programs for machine learning*. San Mateo, California: Morgan Kaufmann.

Rieskamp, J., and U. Hoffrage. 1999. When do people use simple heuristics and how can we tell? In *Simple heuristics that make us smart*, G. Gigerenzer, P. M. Todd, and the ABC Research Group, 141-68. New York: Oxford University Press.

Rumelhart, D. E., J. L. McClelland, and the PDP Research Group. 1986. *Parallel distributed processing: explorations in micro-structures of cognition*. Vols. 1 and 2. Cambridge, Massachusetts: MIT Press.

Schwartz, B., and A. Ward. 2002. Maximizing versus satisficing: happiness is a matter of choice. *Journal of Personality and Social Psychology* 83: 1178-97.

Serwe, S., and C. Frings. 2006. Who will win Wimbledon? The recognition heuristic in predicting sports events. *Journal of Behavioral Decision Making* 19 (4): 321-32.

Shaffer, D., and M. McBeath. 2002. Baseball outfielders maintain a linear optical trajectory when tracking uncatchable fly balls. *Journal of Experimental Psychology: Human Perception and Performance* 28: 335-48.

Shepard, R. 1964. Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology* 1: 54-87.

Simon, H. 1982. *Models of bounded rationality*. Cambridge, Massachusetts: MIT Press.

Smith, V. 2003. Constructivist and ecological rationality in economics. *The American Economic Review* 93: 465-508.

Snook, B., P. Taylor, and C. Bennell. 2004. Geographical profiling: the fast, frugal, and accurate way. *Applied Cognitive Psychology* 18: 105-21.

Snook, B., M. Zito, C. Bennell, and P. Taylor. 2005. On the complexity and accuracy of geographical profiling strategies. *Journal of Quantitative Criminology* 21: 1-25.

Törngren, G., and H. Montgomery. 2004. Worse than chance? Performance and confidence among professionals and laypeople in the stock market. *Journal of Behavioral Finance* 5: 148-53.